改进遗传算法优化 BP 神经网络的语音情感识别

陈 闯[1]，Ryad Chellali[1†]，邢 尹[2]

(1. 南京工业大学 电气工程与控制科学学院，南京 211816; 2. 桂林理工大学 测绘地理信息学院，广西 桂林 541004)

摘 要：语音情感识别技术在人类生活中正扮演着越来越重要的作用。为了更为有效识别语音信号中的情感类型，提出一种改进遗传算法优化 BP 神经网络的识别算法（IAGA-BP）。该算法一方面改进了自适应遗传算法中的选择算子，另一方面更改了自适应遗传算法中的交叉和变异概率公式。通过对自适应遗传算法的改进，提升了遗传算法的寻优性能，并以此对 BP 神经网络初始的权阈值进行优化。在与 BP、GA-BP 和 AGA-BP 网络比较中，实验结果表明，IAGA-BP 网络能够有效提高语音情感识别率，并加快了网路收敛速度。

关键词：遗传算法；BP 神经网络；语音情感识别；自适应；优化

中图分类号：TP183 doi: 10.3969/j.issn.1001-3695.2017.08.0711

Speech emotion recognition based on improved genetic algorithm optimized BP neural network

Chen Chuang[1], Ryad Chellali[1†], Xingyin[2]

(1. *College of Electrical Engineering & Control Science*, *Nanjing Tech University*, *Nanjing* 211816, *China*; 2. *College of Geomatics & Geoinformation*, *Guilin University of Technology*, *Guilin Guangxi* 541004, *China*)

Abstract: Speech emotion recognition technology is playing an increasingly important role in human life. In order to identify the emotion types in speech signals, this paper proposed an improved genetic algorithm optimized BP neural network (IAGA-BP) . This algorithm enhances the optimization performance of genetic algorithm by improving the selection operator and the adaptive crossover and mutation probability formula, such that optimizing the initial weights and thresholds of the BP neural network. The experimental results show that the IAGA-BP network can effectively improve the recognition rate and speed up the convergence compared with the BP, GA-BP and AGA-BP networks.

Key Words: genetic algorithm (GA); BP neural network; speech emotion recognition; adaptive; optimization

0 引言

随着科学技术的日新月异和人类对计算机的依赖性不断增强，人机的交互能力越来越受到研究者的重视。斯坦福大学 Reeves 和 Nass 的研究发现表明[1]，在人机交互中需要解决的问题实际上与人和人交流中的重要因素是一致的，最关键的都是“情感智能”的能力。因此有效识别人类语音信号中的情感类型，分析说话者的情感状态，使得人机交互更为友好，是今后信息处理研究的发展方向 [2]。语音情感识别技术已成功应用于辅助测谎[3]、远程教学[4]、临床医学[5]等各个领域。

用于语音情感识别的方法有很多，主要有高斯混合模型、隐马尔科夫模型、支持向量机和人工神经网络等[6]。人工神经网络是借鉴生物学中神经网络结构和原理提出来的，它模拟了人脑的神经网络，可以处理复杂的输入输出信息，寻找到它们之间的非线性映射关系[7]。其中，BP 神经网络是人工神经网络中应用最为广泛的一种。但由于 BP 神经网络存在易陷入局部极小值、收敛速度慢等问题，而遗传算法具有良好的全局搜索性能，为了进一步提升语音情感的识别率，本文将探讨改进遗传算法和 BP 神经网络相结合的识别模型。

1 BP 神经网络

BP 神经网络是一种包含输入层、隐含层和输出层的不低于 3 层的网络，其结构如图 1 所示。它采用误差函数梯度下降法进行学习，其学习过程主要包括输入信号的正向传播和误差的反向传播过程[8]。在正向传播时，由输入层输入信号，经隐含层处理，传递给输出层。若实际输出与期望输出不符，则转入误差反向传播过程。反向传播时，误差由输出层向输入层逐层传递，并不断修正连接权值，最终使得误差减小。两个过程循环进行，直到均方误差值达到设定标准或达到最大迭代次数为止。但在 BP 神经网络的实际使用中，人们发现它收敛速度较慢、且易陷入局部收敛。其中，网络初始的权值和阈值深刻影响着网络训练结果。

作者简介：陈闯（1992-），男，江苏宿迁人，硕士研究生，主要研究方向为机器学习、语音信号处理；Ryad Chellali（1964-），男（通信作者），法国人，教授，博士，主要研究方向为机器人、虚拟和增强现实、机器学习等（rchellali@njtech.edu.cn）；邢尹（1992-），女，江苏南通人，硕士研究生，主要研究方向为机器学习、GNSS 技术及应用.

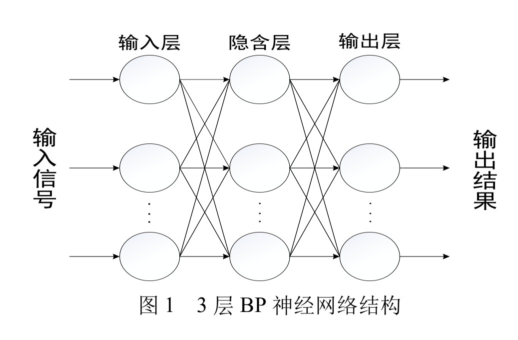


图 1 3 层 BP 神经网络结构

2 改进的自适应遗传算法优化 BP 神经网络

针对 BP 神经网络自身的缺陷，可以利用遗传算法的全局搜索能力来优化 BP 神经网络。遗传算法模拟了生物进化过程，包含选择、交叉和变异等操作，依靠不断地进化，得到最优解[9]。但由于标准遗传算法同样存在易陷入局部最优，影响自身的寻优能力。为此，本文通过改进遗传算法的选择算子以及更改交叉、变异概率公式，以期提升算法寻优性能，进而达到优化 BP 神经网络初始的权阈值目的。

2.1 编码和适应度函数选择

BP 神经网络的权值和阈值为-1 和 1 之间的小数且数量庞大，不适宜二进制编码，这里本文采用实数编码规则。算法搜索过程中，以均方误差的倒数作为评价个体优劣的标准。适应度函数公式如下：

$$fitness(x) = 1/E(x) \tag{1}$$

其中：x 为种群个体，对应于 BP 网络的权值和阈值排列。适应度值越大表示该个体 x 越优秀。

2.2 遗传操作

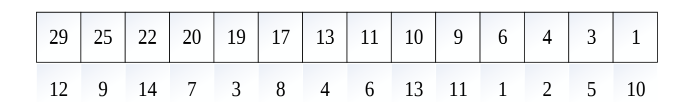
2.2.1 改进选择算子

遗传算法的选择算子常采用适应度比例法和最优个体保存策略。适应度比例法存在随机误差问题，即会出现淘汰适应度高的个体，选中适应度低的个体，体现出不好的竞争力，尤其在大种群数目下更易发生此类问题。最优个体保存策略通过将最优个体直接复制到下一代，避免被交叉和变异操作破坏，但忽略了其余个体破坏种群多样性的可能性[10]。为此，本文提出一种新的选择操作方法，该方法思路新颖，可以在种群中有效地选择较优个体。算法步骤如下：a）确定一个初始种群，计算种群内每个个体的适应度值；b）将种群内的个体按照适应度由大到小降序排列；c）选取前两个个体直接遗传到下一代，其余个体按照已排序好的顺序等分为优、良、中、差四份；d）对优的那份个体复制两份，良、中的个体各复制一份，差的个体不复制。下面以标有个体适应度值和编号的种群数 14 为例，对上述思想进行图形化表示。

a)确定一个种群数为 14 的初始种群：

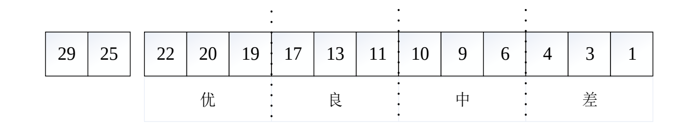
6	4	19	13	3	11	20	17	25	1	9	29	10	22
1	2	3	4	5	6	7	8	9	10	11	12	13	14

b)种群内个体适应度由大到小降序排列：



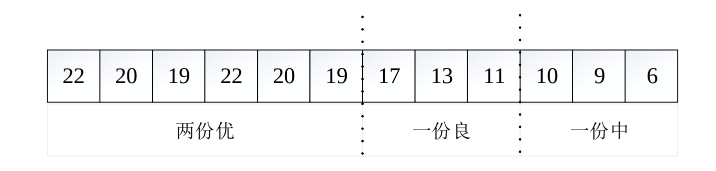
29	25	22	20	19	17	13	11	10	9	6	4	3	1
12	9	14	7	3	8	4	6	13	11	1	2	5	10

c)前两个体直接遗传至下一代，其余个体分为优良中差四份：



29	25	22	20	19	17	13	11	10	9	6	4	3	1
		优			良			中			差		

d)优的个体复制两份，差的个体淘汰，其余个体各复制一份：



22	20	19	22	20	19	17	13	11	10	9	6
两份优						一份良			一份中		

这样选择出的种群，群体平均适应度得到改善，且维持了种群多样性。

2.2.2 交叉操作

根据实数编码规则，从种群中选择两个个体以一定概率交叉得到的新个体，分别为

$$\begin{aligned} w_{kj} &= w_{kj}(1-b) + w_{lj}b \\ w_{lj} &= w_{lj}(1-b) + w_{kj}b \end{aligned} \tag{2}$$

其中：w_{kj} 和 w_{lj} 分别表示第 k 和第 l 个基因的第 j 位；b 为[0,1]的随机数。

2.2.3 变异操作

从种群中随机选择一个个体以一定概率变异得到的新个体为

$$w_{ij} = \begin{cases} w_{ij} + (w_{ij} - w_{max}) f(g), r \ge 0.5; \\ w_{ij} + (w_{min} - w_{ij}) f(g), r < 0.5. \end{cases} \tag{3}$$

$$f(g) = r_2(1 - g / G_{max}) \tag{4}$$

其中：w_{max} 和 w_{min} 分别是基因 w_{ij} 取值的上下界，r 和 r_2 均为[0,1]的随机数，g 为当前迭代次数，G_{max} 为最大进化代数。

2.3 自适应交叉和变异概率

标准遗传算法采用固定的交叉和变异概率，在实际使用中难以调至最佳。当交叉率过小，算法易早熟，交叉率过大，又易破坏优良个体；当变异率过小，算法难产生新的模式结构，变异率过大，又使得算法成为纯粹的随机算法。针对这种问题，Srinivas 等人[11]提出自适应遗传算法（Adaptive Genetic Algorithm，AGA），即种群中个体能够根据周围环境自适应调整交叉和变异概率，具体公式如下：

$$P_c = \begin{cases} \dfrac{k_1(f_{max} - f')}{f_{max} - f_{avg}}, & f' \ge f_{avg} \\ k_2 & f' < f_{avg} \end{cases} \tag{5}$$

$$P_m=\begin{cases}\dfrac{k_3(f_{max}-f)}{f_{max}-f_{avg}}, & f\ge f_{avg}\\ k_4 & f<f_{avg}\end{cases} \tag{6}$$

其中：$f^{'}$ 为待交叉的两个个体中较大的适应度值；f_{max} 为种群的最大适应度；f_{avg} 表示种群的平均适应度；f 表示变异个体的适应度；$k_1\sim k_4$ 均为自适应控制参数。Srinivas 等人认为，当个体适应度低于当代种群平均适应度时，如果该个体被选中则应采用较高的交叉和变异概率，以提升该个体性能；当个体接近当代种群最大适应度时，如果该个体被选中则应采用较低的交叉和变异概率，以尽可能保留该个体的优良模式。但 AGA 在进化初期并不明显，因为此时期群体中的较优个体几乎不发生变化，而此时的优良个体并不一定是全局最优解，这容易使进化走向局部收敛的可能性增加[12]。

2.4　改进的自适应交叉和变异概率

遗传算法中，反映种群适应度值的平均水平用 EX 表示，反映种群适应度值的离散程度用 DX 表示。

$$EX=f_{avg}=\frac{f_1+f_2+\cdots+f_M}{M} \tag{7}$$

$$DX=\frac{f_1^2+f_2^2+\cdots f_M^2}{M}-f_{avg}{}^2 \tag{8}$$

定义相似系数公式[13]为

$$\varphi=\frac{EX+1}{\sqrt{DX}} \tag{9}$$

其中：φ 表示当前种群的相似系数，M 表示种群规模，$f_i(i=1,2,\cdots M)$ 表示种群中个体适应度值。随着遗传算法进化代数的增加，种群适应度平均水平不断提高，重复个体也逐渐增多，使得群体中个体适应度离散程度逐渐减低，并最终使得个体之间越来越相似。依据 AGA 算法的设计原则，结合相似系数定义，改进后的自适应遗传算法交叉概率和变异概率公式为：

$$P_c=\begin{cases}\dfrac{k_1}{1+e^{\frac{k_2}{\varphi}}}\cdot\dfrac{f_{max}-f^{'}}{f_{max}-f_{avg}}+k_3, & f^{'}\ge f_{avg}\\ k_4 & ,\ f^{'}<f_{avg}\end{cases} \tag{10}$$

$$P_m=\begin{cases}\dfrac{k_5}{1+e^{\frac{k_6}{\varphi}}}\cdot\dfrac{f_{max}-f}{f_{max}-f_{avg}}+k_7, & f\ge f_{avg}\\ k_8 & ,\ f<f_{avg}\end{cases} \tag{11}$$

其中：P_c 为交叉概率；P_m 为变异概率；$k_1\sim k_8$ 均为自适应控制参数。当 $f^{'}\ge f_{avg}$ 时，$\dfrac{f_{max}-f^{'}}{f_{max}-f_{avg}}\in[0,1]$；由于 $\dfrac{k_1}{1+e^{\frac{k_2}{\varphi}}}$ 是增函数，且由上分析知相似系数 φ 在不断增大，那么 $\dfrac{k_1}{1+e^{\frac{k_2}{\varphi}}}\in(0,\dfrac{k_1}{2})$；因此，交叉概率 $P_c\in(k_3,\dfrac{k_1}{2}+k_3)$，同理可得变异概率 $P_m\in(k_7,\dfrac{k_5}{2}+k_7)$。本文将自适应交叉概率 P_c 设定于 0.6 和 0.9 之间，变异概率设定于 0.001 和 0.1 之间，k_2 和 k_6 取常数，因此所设定的控制参数 $k_1\sim k_8$ 分别为 $k_1=0.6$，$k_2=1$，$k_3=0.6$，$k_4=0.9$，$k_5=0.198$，$k_6=1$，$k_7=0.001$，$k_8=0.1$。

2.5　IAGA-BP 算法实施步骤

a)给定实验样本，划分为训练样本和测试样本，并进行归一化处理；

b)设置种群规模 N，最大进化代数 G_{max}，初始隐含层节点数等运行参数；

c)随机生成 N 个体作为初始种群，并进行实数编码；

d)对每个个体进行解码，获得隐含层节点的初始连接权值和阈值；训练 BP 神经网络，依据式（1）计算个体适应度值；

e)对个体适应度值进行降序排列，按照 3.2.1 中的改进选择算子选取个体；

f)对选取的个体以式（10）和（11）的自适应交叉和变异概率进行 3.2.2 和 3.2.3 中的交叉和变异操作；

g)判断是否达到最大进化代数，若满足，则输出最优个体；否则，进入 d)；

h)将最优个体解码，获取 BP 网络初始连接权值和阈值，建立 IAGA-BP 模型。

3　基于 IAGA-BP 算法的语音情感识别

本文的语音情感数据库来自于柏林情感语音库[14]。柏林情感语音库由柏林工业大学通过 10 名非专业演员（5 男 5 女）对 10 句录音脚本采用生气、开心、平静、伤心、害怕、厌恶和无聊 7 种情感进行演绎，共录制了 800 条语句。经过 20 名志愿者试听辨别，最终保留了 535 条语句，本实验选取前 5 种语音情感，保留了 400 条语句，训练样本与测试样本按照 1:1 分配。

为有效体现语音信号中的情感信息，本文提取了短时能量、基音频率、共振峰、梅尔频率倒谱系数这 4 类特征及其他们的衍生系数，共构成 140 维的语音情感特征，具体为：短时能量、基音频率、浊音帧差分基音、第一二三共振峰、0-12 阶 MFCC 及一阶差分的最大值、最小值、均值和方差；短时能量的抖动、线性回归系数和线性回归系数的均方误差；0~250Hz 频段能量占总能量的百分比；基音频率的一阶、二阶抖动；第一二三共振峰的一阶抖动；第二共振峰比率的最大值、最小值和均值，具体计算方法详见参考文献[15]。

经过多次实验，选取 140-19-5 的三层 BP 网络结构，具体参数设置为：最大迭代次数 3000，目标精度 0.001，学习率 0.1。GA 参数具体设置为：初始种群规模 50，最大迭代次数 200，交叉率为 0.8，变异率为 0.05。基于柏林情感数据集，分别采用 BP、GA-BP、AGA-BP 和 IAGA-BP 网络进行 5 种情感识别，识别结果如表 1 所示。

从表 1 可以看出，采用 IAGA-BP 网络对 5 种情感的识别率高达 94.50%，其次分别是 AGA-BP 网络的 92.50%，GA-BP 网络的 91.00%和 BP 网络的 90.00%。可以发现，优化后的 BP 网络对于伤心的识别率提升显著，由 BP 网络的 67.74%到 GA-BP 网络的 87.10%和 AGA-BP 网络的 83.87%，尤其是采用 IAGA-BP 网络将伤心识别率提升至 96.77%，说明 IAGA-BP 网络对负面情感的感知更为敏感。一般来说负面情感对人的影响

最大，如果能有效地识别这些负面情感，将有助于提高个体认知和工作效率，减少影响认知和工作能力的因素[16]。在 5 类情感识别中，IAGA-BP 网络对各个情感的识别率分别为 100%、82.35%、94.87%、96.77%、93.94%，除开心情感识别率略低外，其余情感均保持着较高的识别率，说明 IAGA-BP 网络对于语音各类情感识别均衡，具备良好的综合识别性能。

表 1　不同网络的语音情感识别结果

模型	各类情感正确识别率/%					平均识别率/%
	生气	开心	平静	伤心	害怕	
BP	100.00	88.24	92.31	67.74	90.91	90.00
GA-BP	98.41	73.53	92.31	87.10	96.97	91.00
AGA-BP	100.00	76.47	94.87	83.87	100.00	92.50
IAGA-BP	100.00	82.35	94.87	96.77	93.94	94.50

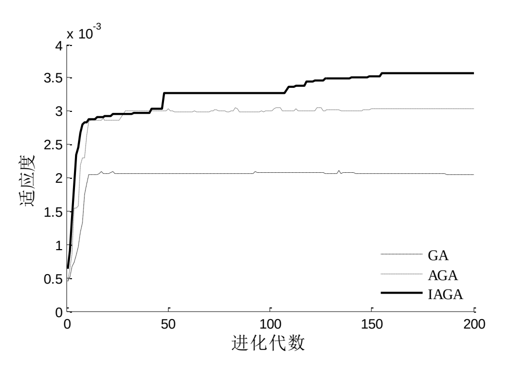


图 2　GA、AGA 和 IAGA 最优个体适应度值曲线

GA、AGA 和 IAGA 三种优化算法的最优个体适应度值曲线，如图 2 所示。由图 2 可见，IAGA 在寻优能力上明显强于 AGA 和 GA。虽然 GA 和 AGA 较快地收敛，但是并没有寻求到最佳的个体。在演化过程中，GA 和 AGA 出现多次适应度下降现象，说明 GA 和 AGA 每次进化过程中，最优个体有被破坏的可能性。IAGA 则通过最优个体保存策略，保护了最优个体；通过分组选择淘汰策略，使得整体水平提高。在 GA 进化至 24 代和 AGA 进化至 29 代时，GA 和 AGA 收敛，都无法逃离旧模式，而 IAGA 则显示了较好的自适应周围环境性能，它通过调整交叉率和变异率逃离旧模式，利用新的模式带动算法继续进行。IAGA 呈阶梯上升状，摆脱了局部收敛，体现出了较强的自适应性能。

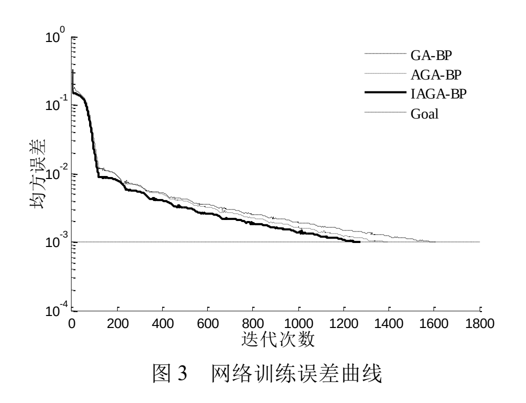


图 3　网络训练误差曲线

GA-BP、AGA-BP 和 IAGA-BP 网络训练误差曲线，如图 3 所示。GA-BP 网络在迭代 1608 次时达到目标精度；AGA-BP 网络在迭代 1392 次达到目标精度；而 IAGA-BP 网络在迭代 1276 次就达到了目标精度，说明 IAGA-BP 网络的收敛速度更快。

4　结束语

本文提出了一种改进的自适应遗传算法优化 BP 神经网络的语音情感识别模型。改进的自适应遗传算法通过个体适应度值大小的排序，选取前两个最优个体直接遗传至下一代，其余个体进行分组选择与淘汰；同时，引入相似系数概念，更改了自适应交叉和变异概率。将改进后的自适应遗传算法应用于 BP 神经网络初始的权阈值优化。在提取柏林数据集 140 维情感特征参数的基础上，分别采用 GA-BP、AGA-BP 和 IAGA-BP 网络对 5 种语音情感进行识别，实验结果表明，采用 IAGA-BP 模型有效提高语音情感识别率，并加快了网络的收敛速度。

参考文献：

[1] Cheng Y M, O'Shaughnessy D, Mermelstein P. Statistical recovery of wideband speech from narrowband speech [J]. IEEE Trans on Speech and Audio Processing, 1994, 2 (4): 544-548.

[2] 韩文静，李海峰，阮华斌，等. 语音情感识别研究进展综述 [J]. 软件学报, 2014, 25 (1): 37-50.

[3] Fragopanagos N, Taylor J G. Emotion recognition in human-computer interaction [J]. Journal of the International Neural Network Society, 2001, 18 (1): 32-80.

[4] Cowie R, Douglas-Cowie E, Tsapatsoulis N, et al. Emotion recognition in human-computer interaction [J]. Neural Networks, 2005, 18 (4): 389-405.

[5] France D J, Shiavi R G, Silverman S, et al. Acoustical properties of speech as indicators of depression and suicidal risk [J]. IEEE Trans on Bio-medical Engineering, 2000, 47 (7): 829-837.

[6] Ayadi M E, Kamel M S, Karray F. Survey on speech emotion recognition: Features, classification schemes, and databases [J]. Pattern Recognition, 2011, 44 (3): 572-587.

[7] Xing H H, Lin H Y. An intelligent method optimizing BP neural network model [J]. Advanced Materials Research, 2013, 607: 2470-2474.

[8] Li X W, Ouyang H B. Evaluating artistic voice of singing objectivly using BPNN [J]. Advanced Materials Research, 2012, 546-547: 1240-1244.

[9] Hou W, Liu X, Chen D, et al. Finance-based scheduling for construction project in China with improved genetic algorithms [J]. Journal of Convergence Information Technology, 2012, 7 (18): 29-38.

[10] 刘浩然，赵翠香，李轩，等. 一种基于改进遗传算法的神经网络优化算法研究 [J]. 仪器仪表学报, 2016, 37 (7): 1573-1580.

[11] Srinivas M, Patnaik L M. Adaptive probabilities of crossover and mutation in genetic algorithms [J]. IEEE Trans on Systems Man and Cybernetics, 2002, 24 (4): 656-667.

[12] 王小平，曹立明. 遗传算法：理论、应用与软件实现 [M]. 西安：西安

交通大学出版社, 2002: 73-74.

[13] Liu Hao, Qiu Jianxin. Improved adaptive genetic algorithm in optimal layout of leather rectangular parts [J]. Advances in Natural Science, 2015, 8 (3): 20-26.

[14] Burkhardt F, Paeschke A, Rolfes M, et al. A database of German emotional speech [C]// Proc of INTERSPEECH. 2005: 1517-1520.

[15] 梁瑞宇, 赵力, 魏昕. 语音信号处理实验教程 [M]. 北京: 机械工业出版社, 2016: 231-234.

[16] 张潇丹, 胡峰, 赵力. 基于改进的蛙跳算法与支持向量机的实用语音情感识别 [J]. 信号处理, 2011, 27 (5): 678-689.